NONLINEAR TIME SERIES ANALYSIS OF MONKEY VOCALIZATIONS

by

Esin Yavuz

B.S. in Computer Engineering, Galatasaray University, 2006

Submitted to the Institute of Biomedical Engineering
in partial fulfillment of the requirements
for the degree of
Master of Science
in
Biomedical Engineering

Boğaziçi University
June, 2008

NONLINEAR TIME SERIES ANALYSIS OF MONKEY VOCALIZATIONS

APPROVED BY:

Assist. Prof. Burak Güçlü (Thesis Advisor)	....
Assoc. Prof. Yasemin Kahya	....
Assist. Prof. Can Yücesoy	....

DATE OF APPROVAL: 24.06.2008

ACKNOWLEDGMENTS

I am grateful to my advisor Dr. Burak Guclu for his help and motivation in finishing my thesis and for encouraging and supporting me during all my studies.

I would also like to thank Dr. Marc Hauser from Harvard University for providing the monkey vocalization database and for his comments about this work.

ABSTRACT

NONLINEAR TIME SERIES ANALYSIS OF MONKEY VOCALIZATIONS

Primate vocalizations are produced as a result of interactions between and within the simple vocal system and the complex signal coming from the nervous system. As a consequence of the nature of this organization, the resulting voice signal is of nonlinear nature. Moreover, in contrast to humans, in many examples of nonhuman primate vocalizations, the vocal folds do not synchronize. Consequently, produced signal is rather complex.

Nonlinear techniques were shown to be useful in analyzing nonhuman primate vocalizations. Deterministic versus stochastic (DVS) prediction technique is one of these methods which can be used to determine the amount of nonlinearity in animal vocalizations. This method serves to calculate the low-dimensional nonlinearity measure (LNM), which indicates the presence of a low-dimensional attractor. By using this method, it was demonstrated that while the nonlinearity measure is useful in voice signals with harmonic components, in highly irregular signals like screams and barks, the detectable amount of nonlinearity was comparatively small.

In this study, the amount of nonlinearity in rhesus monkey voices was calculated by using DVS analysis and this measure was used to distinguish different call types and individual properties of the monkeys. Voice signals with harmonic components showed relatively high SNR and low-dimensional nonlinearity, while these phenomena could not be detected in irregular voices. The signals were analyzed and compared among different callers, different call types and also among call subtypes.

Keywords: Deterministic versus Stochastic Analysis, Monkey Vocalization, Rhesus Macaque, Nonlinearity Measure.

ÖZET

MAYMUN SESLERİNİN LİNEER OLMAYAN ZAMAN SERİSİ ANALİZİ

Primat sesleri, mekanik ses üretim sistemi ile sinir sisteminden gelen karmaşık sinir işaretinin etkileşiminin bir sonucudur. Bu sistemin doğası gereği, sistemin oluşturduğu ses işareti nonlineerdir. İnsan olmayan primatların seslerinde görülen nonlineerliğin ses üretim sisteminin yapısal özelliklerinden kaynaklandığı ve karmaşık bir sinirsel kontrol mekanizmasının varlığını zorunlu kılmadığı öne sürülmüştür.

Hayvan seslerindeki nonlineerlik miktarı DVS öngörü yöntemiyle belirlenebilir. Bu yöntem, sistemde az-boyutlu bir çekicinin bulunduğunu gösteren LNM değerini hesaplamaya yarar. Bu yöntem kullanılarak, harmonik bileşenler içeren seslerde bu ölçütün iyi sonuçlar vermesine rağmen düzensiz seslerde tespit edilebilen nonlineerliğin nispeten düşük olduğu gösterilmiştir.

Bu çalışmada rhesus makaklarının normal ve agresif tip ses işaretlerindeki nonlineerliği tespit etmek ve anlamlandırmak için bir nonlineer zaman serisi analizi yöntemi olan DVS analiz yöntemi kullanılmıştır. Harmonik bileşenler içeren seslerde az sayıda komşu kullanılarak yapılan öngörülerin daha iyi sonuç verdiği; buna karşın düzensiz seslerde komşuluk sayısının öngörü başarısında bir fark yaratmadığı gözlemlendi. Sesler incelendikten sonra farklı bireyler, farklı ses tipleri ve ses alt grupları ayırt edilmeye çalışıldı.

Anahtar Sözcükler: DVS Çözümleme, Maymun sesi, Rhesus Makakı, Nonlineerlik Ölçütü

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

d	Embedding Dimension
τ	Delay Time for Embedding
D	Number of Neighbors for Prediction

LIST OF ABBREVIATIONS

LNM	Low-dimensional Nonlinearity Measure
DVS	Deterministic vs. Stochastic
SNR	Signal to Noise Ratio
AMI	Average Mutual Information

1. INTRODUCTION

1.1 Motivation and Objectives

Voice is the most important medium for primate communication. By shaping the air flow by larynx, vocal tract and respiratory cavities, not only humans, but also other primates produce vocal signals that correspond to a large diversity of social context [1].

Nonlinear phenomena have been shown in human [2] and nonhuman mammal [3, 4] voices. Moreover, these works showed that while humans avoid producing irregular voices by using the vocal folds as coupled oscillators; this is not the case in nonhuman mammals. The morphological features of vocal folds and laryngeal muscles in nonhuman primates indicate that in contrast to the human larynx, the nonhuman primate larynx allows for a greater range in vocal pitch as well as greater instability [5]. Furthermore, nonlinearities are common in nonhuman mammal vocalizations and this may play a role in their communication: the voice signal may carry some information about the properties like age, size, and mood of the animal.

Common structures and behaviors that are seen in nonlinear systems such as steady state attractors, limit cycles, subharmonics, chaos and bifurcations can be distinguished form the time series data of mammal vocalizations [4]. These phenomena can also be observed in spectrographic analysis of voice signals. Furthermore, computer simulations of biomechanical modeling showed that these irregularities may be resulted from the desynchronization of the left and right vocal fold or due to the desynchronization of vertical and horizontal vibratory modes of a single fold [6, 7, 8].

Deterministic versus Stochastic (DVS) prediction technique was shown to be efficient for investigating the nonlinear phenomena by quantifying the amount of nonlinearity in animal vocalizations [9]. In the DVS technique, deterministic and stochastic prediction efficiencies are compared and as an indication of this difference, low-dimensional nonlinearity measure (LNM) is calculated. A high LNM is found when deterministic prediction is more efficient than stochastic prediction and it indicates the presence of a low-dimensional attractor.

Low-dimensionality of a system may be a result of synchronization of several components in order to produce harmonic components. Furthermore, it has been demonstrated by Ruelle and Takens that in a system with many degrees of freedom, when there is a transition to chaos, many degrees of freedom are coupled and the number of dimensions drastically decreases [10]. After the transition to chaos, the number of degrees of freedom may increase. Nevertheless, DVS technique gives good results about the low-dimensionality of the system even if the signal shows partial low-dimensional behavior.

A usual way to investigate a nonlinear system is to model the system by using differential equations and analyzing the results. However, the vocal system consists of many components and it is very difficult to estimate the dimensions of the attractor by numerically integrating the partial differential equations. Instead dealing with this problem, DVS analysis uses the state space reconstruction technique [11]. The main question here is whether the time series from a high dimensional deterministic system can be approximately modeled with a low dimensional non-linear stochastic model, as has been suggested when there are large, spatially coherent structures in the system [12]. For a high dimensional nonlinear system with low level observational noise, a large noise term can be induced in the prediction step by the process of state space reconstruction from time series data, consequently it is not possible to make accurate short-term forecasting of the time series, irrespective of the length of the time series [13]. As a result of that, a high dimensional nonlinear system is equivalent to a stochastic system, which results in a low LNM.

Tokuda et al. [9] used the DVS method to show that while the nonlinearity measure is useful in voice signals with harmonic components, in highly irregular signals such as juvenile macaque screams, piglet screams, and some dog barks, the detectable amount of nonlinearity was comparatively small. It has been discussed that the nonlinearities in nonhuman primate vocalizations may be a consequence of the interactions between structural properties of the peripheral production mechanism, which allows individuals to generate highly complex and unpredictable vocalizations without requiring a complex neural control mechanism [3].

Analyzing nonhuman primate vocal communication is indispensable for investigating the evolution of speech and language. In addition, understanding the constraints on the perceptual and motor domains of primates' vocal behavior will be a big step to understand their cognitive abilities [14].

In the present study, amount of nonlinearity of monkey voices was determined by DVS method and individual properties of monkeys and social contexts were compared with respect to the nonlinearity measure.

1.2 Outline

The work is presented as follows: In Chapter 2, background information about the anatomy and the physiology of vocal system is given. In Chapter 3, materials and methods used in the present study are explained. Next, results are given in Chapter 4. Finally, results are discussed in Chapter 5 and conclusion is made in Chapter 6.

2. VOICE PRODUCTION MECHANISMS IN HUMAN AND NONHUMAN PRIMATES

2.1 Comparison of Human and Nonhuman Vocal Systems

Voice signal is produced by shaping the air pumped by the lungs in vocal tract. Even if there are some important differences; vocal system anatomy, especially the vocal production mechanism, is very similar in humans and nonhuman primates.

The main source for voice production is the air pressure coming from the lungs. The air pressure is maintained and adjusted by the lungs, diaphragm, chest and abdominal muscles.

Air flow coming from the respiratory system then passes the larynx and vocal folds. The main function of the larynx is to choose between swallowing and breathing actions and thus protect the lungs. Its function in voice production is to regulate vocal fold tension. Nonhuman primates also have air sacs attached to their larynges. These large sacs do not exist in humans and their functions are not clearly known.

The basic sound of the voice is produced by the vibration of vocal folds. The rate of vibration is called the vocal pitch. Vibration occurs in vibratory cycles. At first, air pressure opens the bottom of vocal folds. Then, the pressure moves upwards through the vocal folds and opens the top. After this step, bottom of the vocal folds is closed because of the "Bernoulli effect" created by the low pressure behind the fast air movement. This causes the vocal folds close and cut the air flow, and release an air pulse [15].

From physical point of view, vocal folds can be thought as coupled oscillators. This coupled oscillator system is the main structure that creates the nonlinear phenomena in voice.

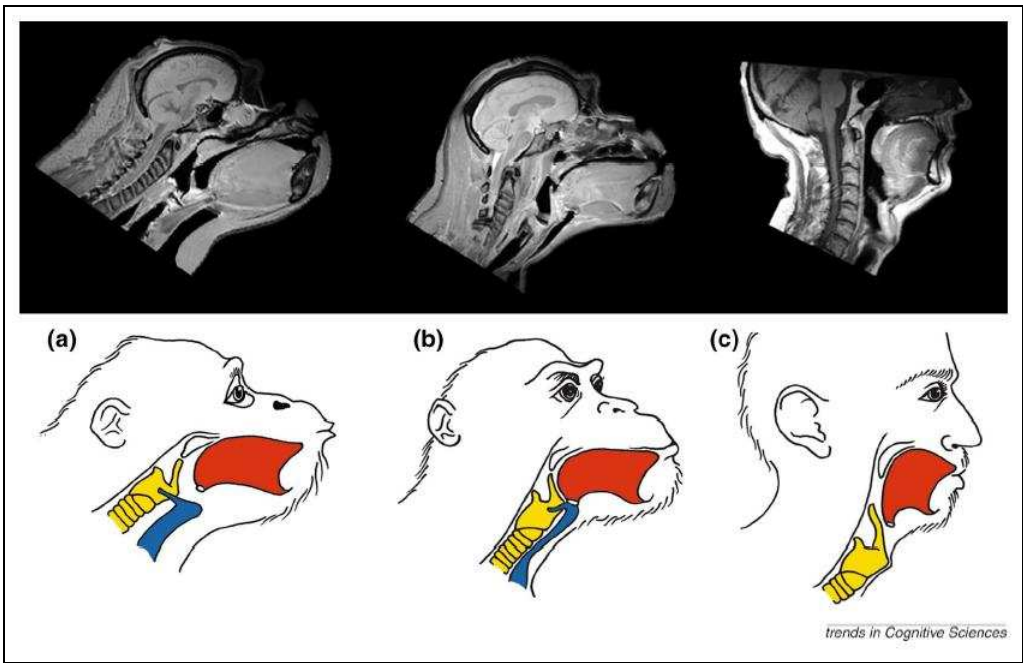


Figure 2.1 Anatomy of voice production in orangutan, chimpanzee and human (a to c, respectively). Upper panel shows MRI images of vocal system, and lower panel contains the illustrations of the vocal system of human and other primates. Red color signifies the tongue body, yellow the larynx and blue the air sacs [1]

After the voice sound is produced by the vibrations of vocal folds, it is finally shaped in vocal tract, which consists of the throat, nasal and oral cavities. This shaping is realized by a series of band-pass filters called *formants*. This filtering process is independent of the vocal pitch determined in the vocal folds. Therefore, larynx and vocal tract can be thought as two independent systems. This theory of voice production is called the *source/filter theory*. The acoustics, anatomy, innervation and central control of human and animal vocal tracts are fundamentally similar [1]. For humans, individual recognition is a result of this shaping, namely the *resonance*.

The final step, *articulation*, is provided by shaping the sound by lips, tongue and soft palate in order to produce more specific vocalizations, which correspond to the words in human vocalization. This step is very weak in nonhuman primates, therefore they can not develop languages as humans do. Nevertheless, Hauser et al. showed that rhesus monkeys have the ability to modify the spectral characteristics of the signal by modifying the lip protrusion, lip separation, teeth separation and mandibular position, and maybe also the tongue position [16], and this modifications may play a role in producing distinct sounds.

Comparison of primate vocal systems is shown in Figure 2.1. The figure shows that the human larynx is placed lower in the throat than in the apes. Moreover, oral cavity is longer and tongue shape is different in humans compared to the other primates. These differences allow a much greater range of sounds to be produced by humans, which would have been significant in the evolution of speech [1].

3. METHODOLOGY

3.1 Monkey Voice Database

The data used in this study includes digitized vocalizations of rhesus macaques obtained from Harvard University Primate Cognitive Neuroscience Laboratory. The database contains monkey calls that are recorded from rhesus monkeys living on the island of Cayo Santiago, Puerto Rico by Marc Hauser and field assistants.

The database contained examplars from 10 major call types. Some of these call types contained some additional call subtypes. Within these call types, coos, aggressive calls, girneys, grunts, harmonic arches, and screams are examined.

The database included different call type exemplars from different individuals. For some individuals, multiple examplars of the same call type were also provided. All voice signals used in this study were recorded from adults; on Cayo Santiago Island, females reach reproductive maturity at approximately 3 years and males at approximately 4 years.

The database included different call types, six of which are analyzed here. These calls differ in context and in acoustical properties:

Coos: Coos are calls that are produced during a variety of social interactions like friendly approaches or approaching a common food. It has been reported that individuals produce very distinctive coos [5].

Aggressive calls: These calls are the vocalizations that are produced as a threat or in a fight. The sound database included three types of aggressive calls: pant threats, barks and growls. All signals were recorded from adults.

Girneys: Girneys are the voice signals that are produced during social interactions such as grooming and handling of infants by females. The characteristic of this signal type is the drop of fundamental frequency over the course of the bout.

Screams: These signals are produced when the animal is under threat or attack of another dominant animal. Five subclasses of screams have been reported [17]: tonal, noisy, arched, pulsed and undulating. The subgroups differ in frequency and complexity, and each subtype is given in a different context. Noisy screams were produced in order to call help by juveniles when a higher rank animal attacked. Undulating screams also told of an attack by a higher ranking opponent, but without physical contact. Arched screams indicated a lower ranking aggressor and did not indicate any physical contact. Pulsed and tonal screams tended to indicate a squabble within the immediate family. The screams are thought to carry information about the situation, the location and identity of the calling individual in addition to the degree of fear. It is important to note that although the monkeys appeared to distinguish the sounds easily, human researchers had to rely on voice prints at first.

Grunts: these signals are given during social interactions like approaching to groom, approaching to common food items, and group movement. They sound like pant threat type of aggressive calls, but in contrast to pant threats, they are given in a friendly context.

Harmonic Arches: These are given when a high quality, rare food is discovered.

3.2 Preprocessing of Signals

Before starting the nonlinearity analysis, all the samples were normalized in order to make them have the same properties.

In the original database, signals were provided with different sampling rates. First, all signals were downsampled to 20020 Hz. After resampling, the amplitudes were normalized to 1 V_{RMS} . Silent periods at the beginning and at the end of the recordings were discarded.

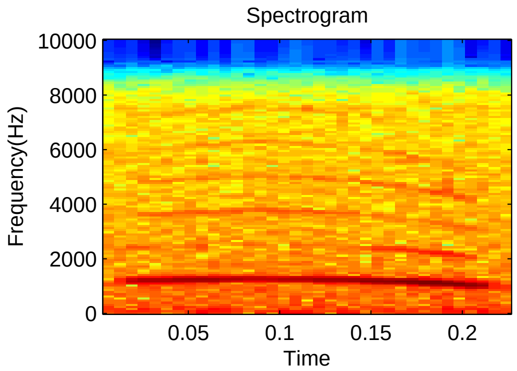
Frequency range of each signal was determined by power spectral density estimation. Spectrograms of some samples are shown in Figure 3.1. Harmonic components can be clearly seen in coos and harmonic arches while the frequency components of aggressive calls are not very clear. Screams also have clear harmonic components, but in this case the signal is not stationary.

3.3 Nonlinear Time Series Analysis

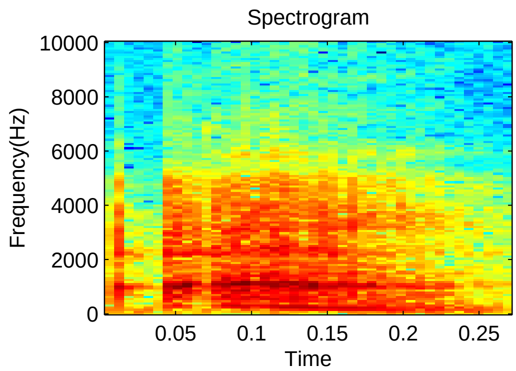
A time series is a discrete time sequence of data observed from one or more channels from a system. Time series analysis methods are widely used in many different research areas as they serve to extract information about the underlying mechanisms.

Most of the time series analysis methods are linear prediction models but nonlinear models have also been introduced. Time series analysis is a very important tool for investigating the systems that exhibit nonlinear dynamics since we generally do not know the exact components of the system. The only data we have is usually the one dimensional output of the system, namely the time series data. Nonlinear time series analysis methods are very useful to construct the original phase space of the system and investigate the underlying dynamical behavior.

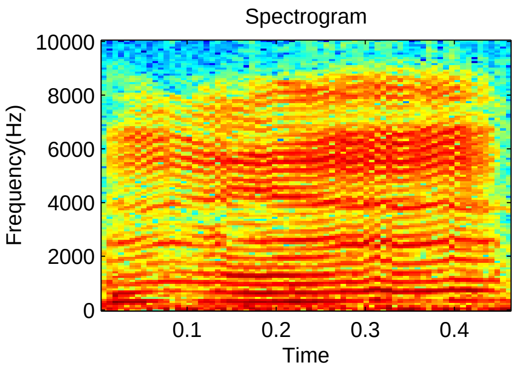
Casdagli [13] introduced a method that combines and compares a deterministic nonlinear prediction model [18] with a stochastic linear prediction model [19] by means of prediction accuracy. If a low-dimensional attractor exists at least partially in the system, the deterministic model would give more accurate prediction results than the stochastic model. The difference of prediction accuracy between the linear and nonlinear prediction models, namely the low-dimensional nonlinearity measure (LNM) estimates the strength of nonlinearity in the signal.



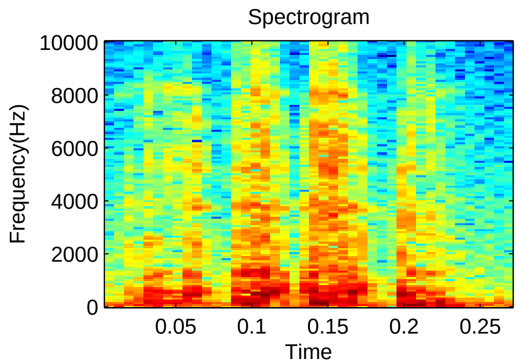
(a) Coo



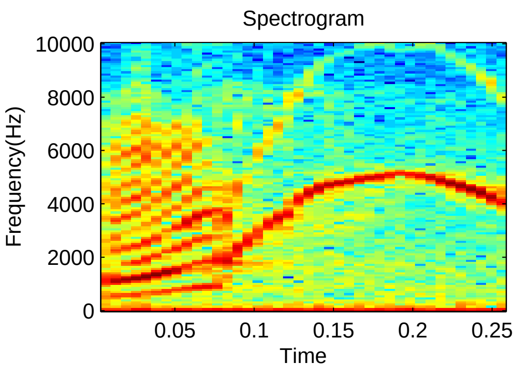
(b) Aggressive Call



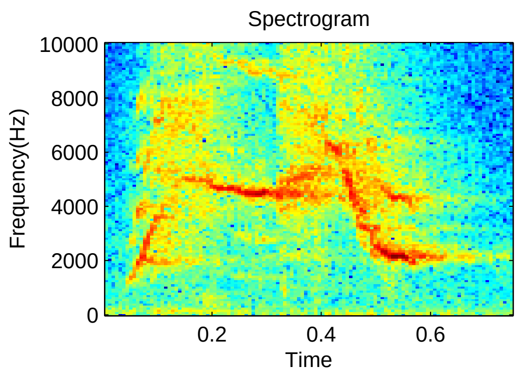
(c) Girney



(d) Grunt



(e) Harmonic Arch



(f) Scream

Figure 3.1 Spectrograms for exemplars from each call type

3.3.1 Deterministic Versus Stochastic Analysis

DVS analysis is based on Takens embedding theorem [11] and compares prediction success of deterministic and stochastic prediction methods.

According to this theorem, embedding is done as follows:. Each vector in the delay coordinate space consists of the data point at time t, and d data points before that point. Data points in the vector are τ elements far from each other.

$$x(t) = (x_t, x_{t-\tau}, x_{t-2\tau}, ..., x_{t-(d-1)\tau}) \tag{3.1}$$

where d refers to the embedding dimension and τ is the delay time.

After that, for every data point $x(t)$, distances from other points are calculated and D closest neighbors of $x(t)$ are found. The data point itself and its temporally close points are not included in the neighbors. Then, one step further state of x(t) is predicted by using a local linear predictor as in Eq. 3.2.

$$\tilde{x}_{t+1} = \sum_{k=0}^{d-1} a_k(t) x_{t-k\tau} \tag{3.2}$$

In Eq. 3.2, the prediction coefficients $a_0(t), a_1(t), ..., a_{d-1}(t)$ are determined by a Least-Square algorithm for D neighbors. The prediction accuracy is computed by finding the difference between the predicted signal and the actual signal. This difference gave the residual signal (r). The signal to noise ratio (SNR) is calculated as in Eq. 3.3.

$$SNR[dB] = 10log\left[\frac{\sum_{t=d}^{N}\{x_t - \bar{x}\}}{\sum_{t=d}^{N}\{r_t - \bar{r}\}}\right] \tag{3.3}$$

where

$$\bar{x} = \frac{1}{N-d+1}\sum_{t=d}^{N} x_t \ , \ \bar{r}\frac{1}{N-d+1}\sum_{t=d}^{N} r_t. \tag{3.4}$$

SNR is calculated for different number of neighbors by increasing the percentage of the number of neighbors from 0 to 100. For very small number of neighbors, the prediction is very sensitive to noise; therefore the SNR is very low. As the number of neighbors increases, the SNR also increases. For a nonlinear dynamical system, the SNR achieves an optimum value for an intermediate number of neighbors. As the number of neighbors is further increased, linear prediction does not give accurate results because of the nonlinear nature of the system. Global linear prediction which uses the maximum number of neighbors is almost identical to AR modeling [9]. The difference between the optimum SNR and the global-linear-prediction SNR gave the LNM.

Examples of DVS analysis are applied to simulated sine wave, Lorenz system, and random data signals. 3D projection of the phase space and SNR are shown in Figure 3.2.

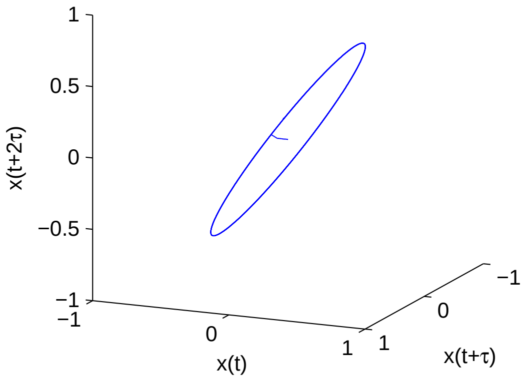
Lorenz data is simulated by the Runge-Kutta integration with time step 0.01. The Equation used in simulation is given in Eq. 3.5.

$$\begin{aligned}\frac{dx}{dt} &= \sigma(y-x)\\ \frac{dy}{dt} &= -xz+rx-y\\ \frac{dz}{dt} &= xy-bz. \end{aligned} \tag{3.5}$$

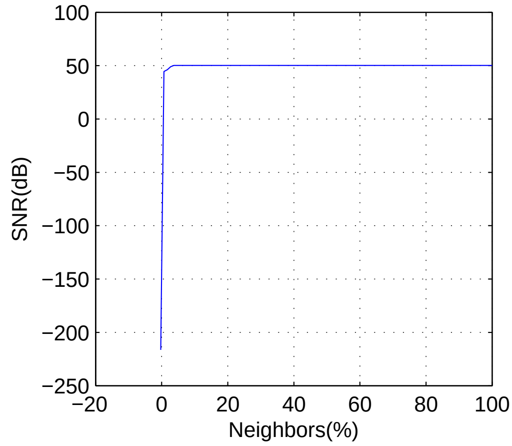
where σ=16.0, b=4 and r=45.92. With these parameters, the system shows chaotic behavior.

Sine wave is generated between -5π and 5π with time step 0.01, and random data is generated by *rand* command of MATLAB.

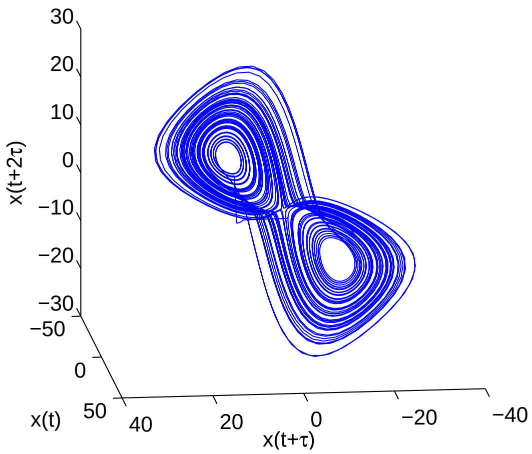
In Figure 3.2, it can be seen that the phase space reconstruction can generate the Lorenz attractor and also the LNM measure is high for the Lorenz system. For the sine wave, the limit cycle can be reconstructed by embedding, and the LNM measure is very low as the sine wave can be obtained by linear models, and therefore stochastic



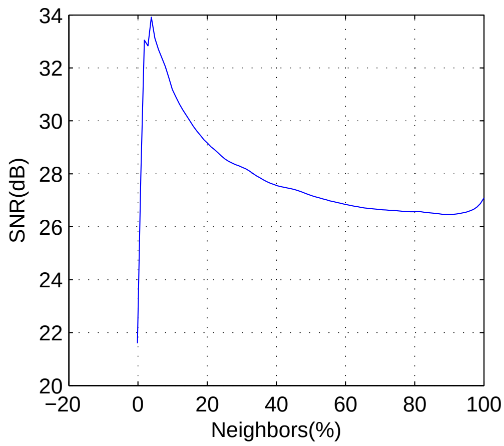
(a) Sine Wave phase space



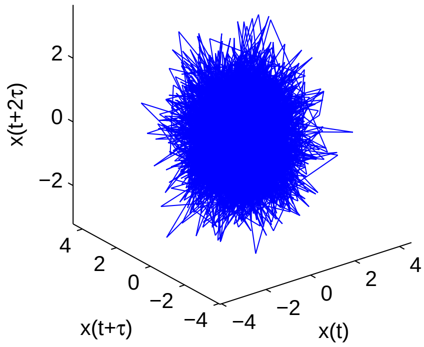
(b) Sine Wave SNR



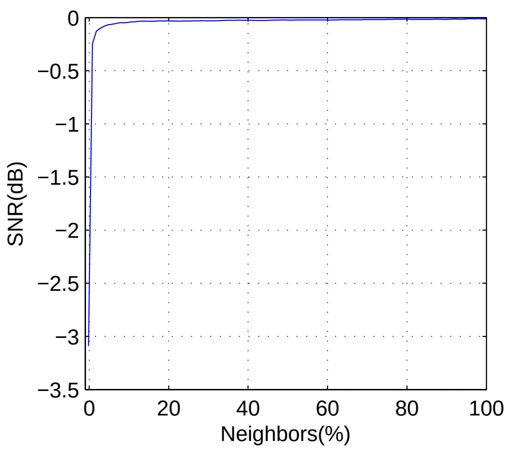
(c) Lorenz System phase space



(d) Lorenz System SNR



(e) Random Data phase space



(f) Random Data SNR

Figure 3.2 Application of DVS method to computer data prediction gives results that can be obtained as good as by using deterministic models. Finally, no particular shape can be obtained in the 3D projection of the phase space of random data, for which the LNM is very low. One important point here is that both random data and sine wave shows low LNM; however the amount of SNR is high for sine wave while it is low for random data.

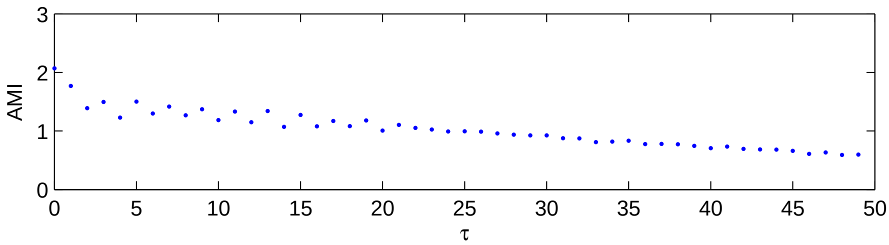
3.3.2 Parameter Estimation

Choosing the time delay is a critical factor for DVS analysis. If the time delay is too small, it is not possible to observe a significant change in one time step. On the other hand, if the time delay is chosen to be too large, data points will be uncorrelated; especially if the system is chaotic; because in a chaotic system, nearby trajectories diverge exponentially fast. The correlation between two points in a delay coordinate space can be estimated by finding the average mutual information between these two points. Average mutual information (AMI) gives the amount of information available for a point by making an observation at another point. The AMI between the observations [20] at two different points, $s(n)$ and $s(n+\tau)$ is

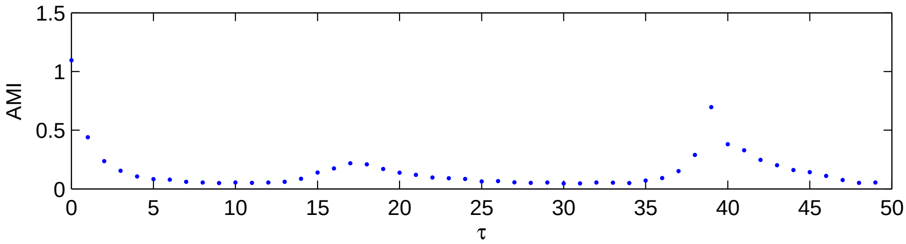
$$I(\tau) = \sum_{n=1}^{N} P(s(n), s(n+\tau)) \left[\log_2 \frac{P(s(n), s(n+\tau))}{P(s(n)P(s(n+\tau))} \right] \tag{3.6}$$

If the time delay is too small, the system may not evolve enough in one time step, therefore $s(n)$ and $s(n+\tau)$ will be correlated, which leads to a high value of AMI. On the other hand, if the time delay is too high, two consecutive signals may be uncorrelated, especially if the system is chaotic. Therefore, the first minimum of $I(\tau)$ gives an appropriate choice for the time delay.

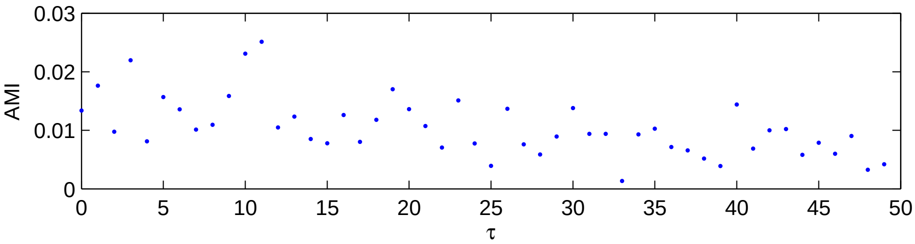
Examples of AMI can be seen in Figure 3.3. These figures correspond to the AMI of the sine wave, Lorenz system, and the random data of which LNM measures and 3D phase space peojections are shown in Figure 3.2.



(a) Sine Wave



(b) Lorenz System



(c) Random Data

Figure 3.3 Results for AMI for the computer generated signals

4. RESULTS

DVS method is applied to voice signals from six call types of all callers. The number of exemplars from each call type is given in Table 4.1.

Before starting the analysis, delay time parameter for DVS analysis was calculated. Delay time τ was determined by AMI method. Best time delays found by this method are shown in Table 4.2 for each call.

Table 4.1
Number of samples from each call type

Call Type	Number of Samples
Coos	25
Aggressive Calls	54
Pant Threats	9
Growls	7
Barks	34
Other	4
Screams	32
Arched	5
Tonal	5
Pulsed	4
Noisy	2
Other	16
Girneys	16
Harmonic Arches	16
Grunts	47

To determine the optimal delay time for embedding, AMI for every signal wss calculated. The average of the first minima for each group was found and this number was chosen as the delay time parameter for DVS analysis.

Embedding theorem states that embedding can succesfully be done in any dimension bigger than two times the original dimension of the system. In this study, embedding dimension was chosen to be d=10 for all the signals, as a dynamical system can be considered low dimensional if it exhibits a few (approximately smaller than 10) degrees of freedom [9].

Table 4.2
Time delay parameters for DVS Analysis

Call Type	Delay Time
Coos	8
Aggressive Calls	8
Screams	5
Girneys	7
Harmonic Arches	4
Grunts	6

Time and frequency domain representations and parameter estimation results for a coo call which has a high LNM are shown in Figure 4.1, and the results for an aggressive call which has low LNM are represented in Figure 4.2. The Spectrogram of the coo call reveals the fundamental frequency and its harmonics. By embedding the data in delay coordinate state space, it was possible to reconstruct the limit cycle to which the trajectory converges. This limit cycle corresponds to the fundamental frequency. Conversely, the aggressive call exemplar is very irregular and it is not possible to extract any information about the existence of a fundamental frequency from the spectrogram. For this irregular signal, no particular attractor could be obtained by embedding the time-series data in delay coordinate state space, hence the data can be thought as output of a stochastic process. Thus, LNM measure was very low. State space reconstructions and SNR as a function of dimension percentage for these two signals are compared in Figure 4.3.

Using the LNM measures, at first different call types from all animals were analyzed and difference between call types were compared.

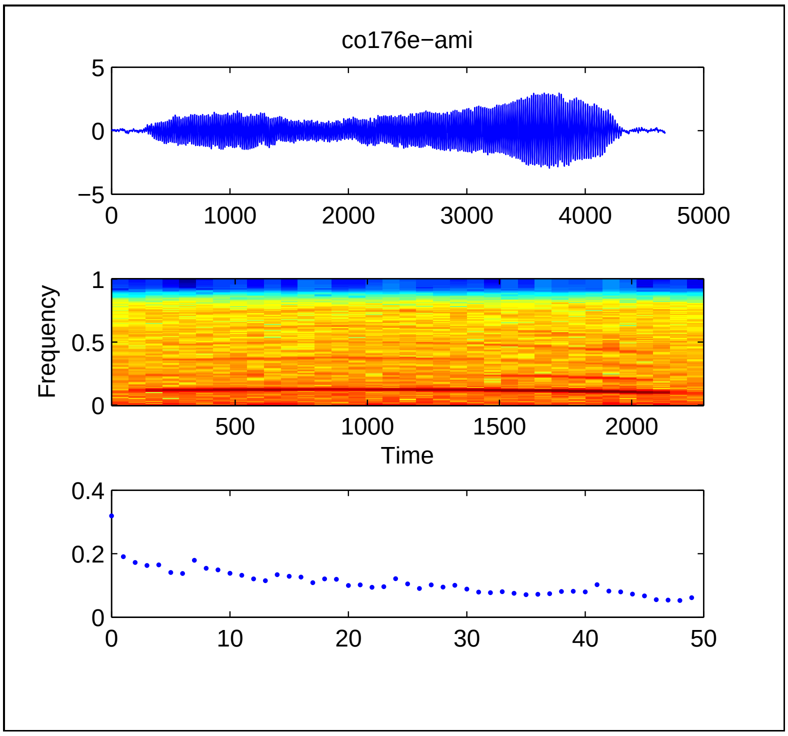


Figure 4.1 Preprocessing and parameter estimation for a coo call. In descending order, panels correspond to the time-series data (normalized amplitude vs sample number), normalized Fourier spectrogram (normalized frequency vs. time), average mutual information (AMI vs. τ)

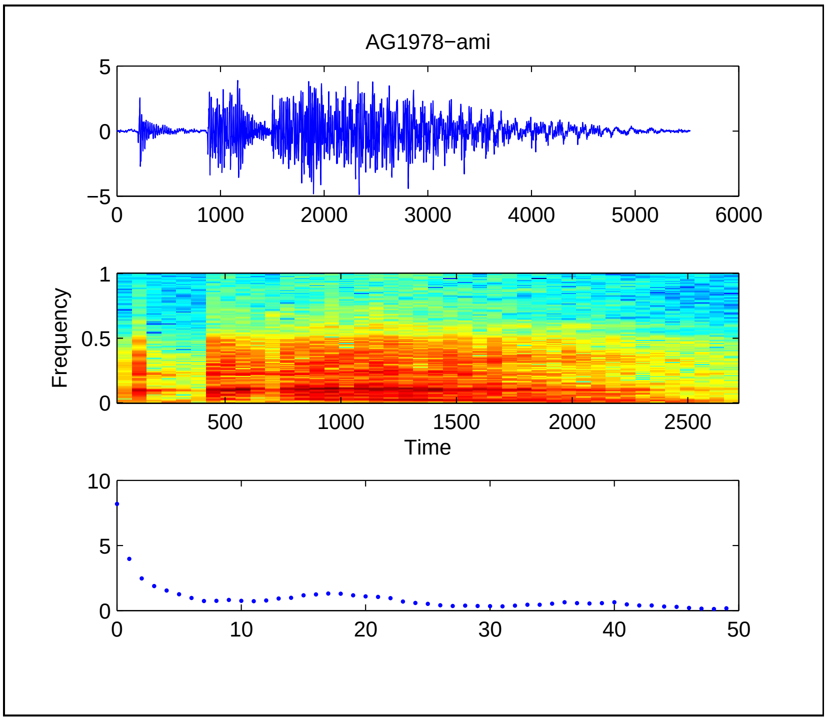


Figure 4.2 Preprocessing and parameter estimation for an aggressive call. In descending order, panels correspond to the time-series data (normalized amplitude vs sample number), normalized Fourier spectrogram (normalized frequency vs. time), average mutual information (AMI vs. τ)

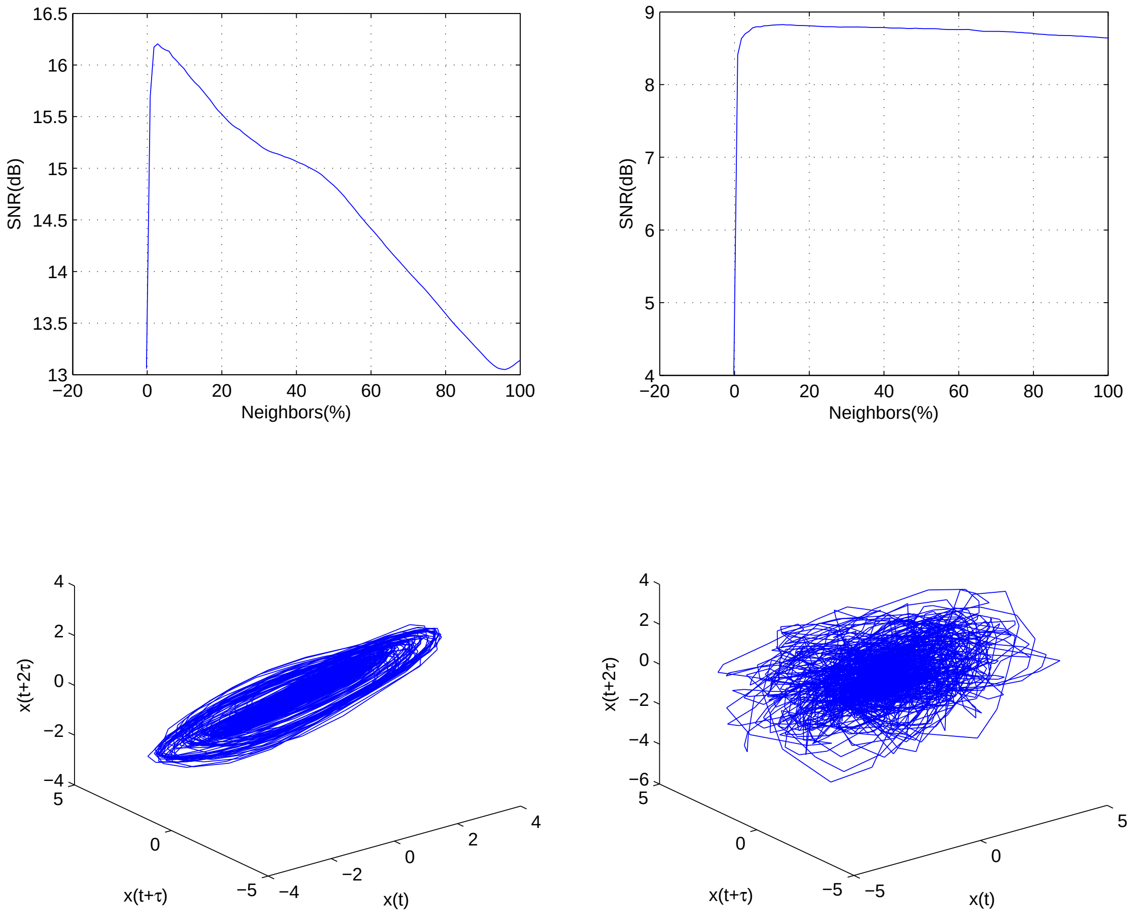


Figure 4.3 3-D projections and SNR vs. neighborhood percentage for the signals in Figures 4.1 and 4.2. Difference between maximum SNR and final SNR gives the LNM.

After that, coo signals from different individuals were compared and the difference among callers was examined.

Next, different call types from same caller were analyzed and nonlinearity difference between call types of same caller was identified.

To human ear, grunts are very similar to pant threat type aggressive calls. In order to see if the nonlinearity measure can distinguish this difference, pant threat LNM mean was compared to grunt LNM mean.

Aggressive calls and screams in the dataset were divided into subgroups. Finally, nonlinearity difference among aggressive and scream subgroups was analyzed.

4.1 Comparison of Different Call Types

LNM values of 6 call types were compared by unbalanced one-way ANOVA test. Resulting p-value is nearly 0 which means that the null hypothesis that claims that the call types has the same mean estimate can be rejected. Therefore this test shows that the difference between call groups is highly significant.

Resulting statistics were analyzed by a multivariate data analysis with Tukey-Kramer criterion. This test showed that coo mean is significantly different from means of other five call types, while means of grunts and aggressives are similar and also scream, harmonic arch and girney means are not significantly different from each other. MATLAB output of differences among call types is shown in Figure 4.4.

Another phenomenon that can be observed in Figure 4.4 is that coo mean is higher that other call type means, and the grunt has lowest mean among all call types. Mean LNM estimates, standard deviation and mean of maximum SNR are given in Table 4.3. This table shows that for coos, not only LNM mean, but also maximum SNR mean is also far higher than other call types.

Table 4.3
Comparison of LNM and SNR of call types

Call Type	Mean LNM	Standard Deviation	Mean Maximum SNR
Coos	3.6455	1.6711	17.0468
Aggressive Calls	0.1105	0.1395	8.38
Screams	2.1132	1.6846	7.103
Girneys	0.6718	0.4600	10.88
Harmonic Arches	2.8021	1.4948	9.1231
Grunts	0.0532	0.0573	9.2089

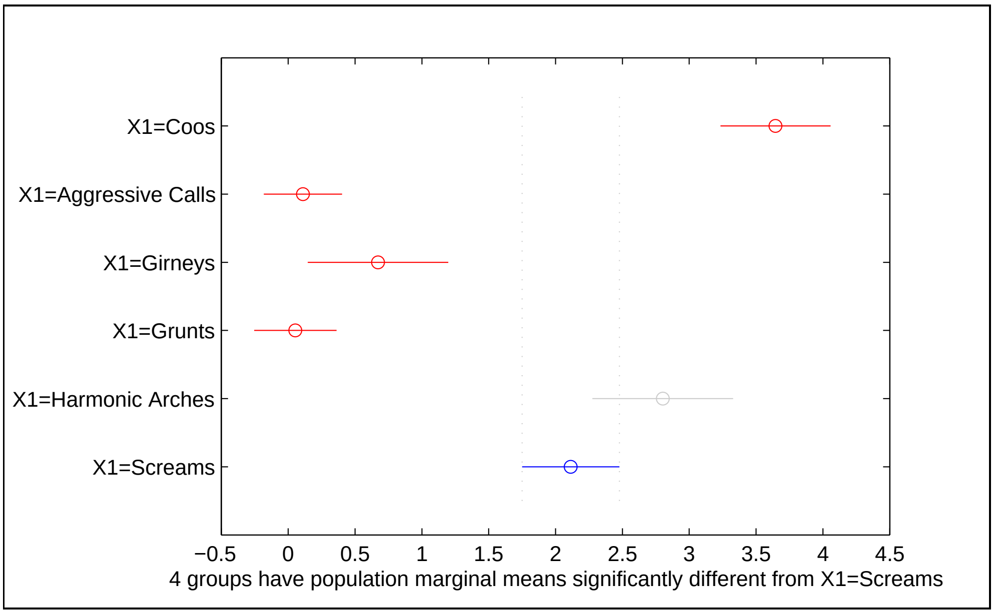


Figure 4.4 Difference between call types. Mean of coo calls is significantly different from means of other calls except harmonic arches. Mean of aggressive calls is close to the mean of grunt and girney calls. Mean of harmonic arches is close to the mean of screams.

4.2 Coo Calls among Different Callers

In order to discriminate individuals by the nonlinearity measure, DVS method is applied to coo calls of two individuals (n=4 from one, n=6 from the other, and LNM measures were analyzed by two-sample t-test. Resulting p-value is 0.47 and therefore t-test does not reveal any significant difference between the nonlinearity of coo samples from these two animals.

4.3 Different Type Calls of an Individual

Different type calls of one animal were analyzed (n=6 for coos, n=5 for girneys, n=2 for harmonic arches, n=7 for grunts) by unbalanced one-way ANOVA method. Resulting statistics were tested by multivariate comparison test. Comparison of means of different call types or a single individual calculated by this analysis is shown in Figure 4.5. It was possible to test only four call types as no signals from other types were provided for this individual. Mean LNM, standard deviation of the LNM and

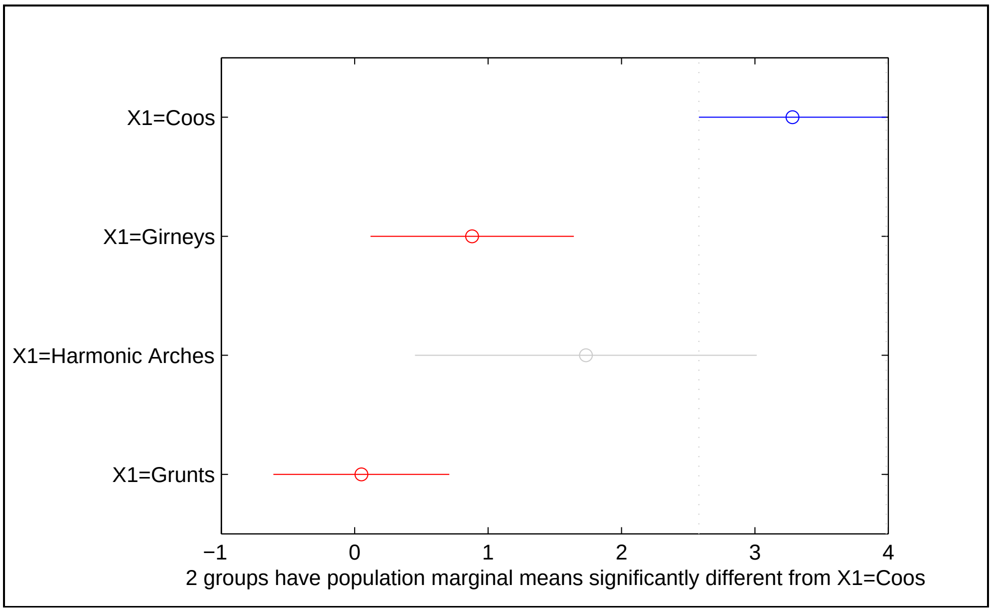


Figure 4.5 Comparison of means from different call types. Mean of coo calls is significantly different from means of girney and grunt type calls.

mean of the maximum SNR obtained are given in Table 4.4

Table 4.4
Mean and standard deviation of LNM, and mean of the maximum SNR for different call types of an individual

Call Type	Mean LNM	Standard Deviation	Mean Maximum SNR
Coos	3.2822	1.4385	11.6083
Girneys	0.8804	0.5431	8.2440
Harmonic Arches	1.7328	0.1314	4.7
Grunts	0.0498	0.0139	11.3871

The analysis shows that the difference of nonlinearity between call types for a single animal is also significant as for the analysis for all animals in the first test ($p < 0.05$). Again, highest mean among call types is obtained for coos.

4.4 Pant Threats vs. Grunts

Grunt and Pant threat type aggressive calls were analyzed using two-sample t-test (n=9 for pant threats, n=47 for grunts). Results show that there is no significant difference between these call types (p=0.583).

4.5 Subgroups Analysis

Aggressive calls and screams were divided in subgroups by the contexts in which they are given. LNM measures were compared among subgroups for both call types separately.

4.5.1 Aggressive Subtypes

The dataset included 3 types of aggressive calls: Pant threats (n=9), growls (n=7) and barks (n=34). Nonlinearity measures of these subgroups were compared by unbalanced one-way ANOVA test. This analysis revealed that the difference between subgroup means is not significant (p=0.421).

Resulting statistics were examined by multivariate data analysis. Comparison of means is represented in Figure 4.6.

Figure 4.6 shows that none of the subgroup's mean is significantly different from other subgroups.

4.5.2 Scream Subtypes

Scream subgroups provided in the database were analyzed and LNM measures per subgroup (n=5 for arched, n=2 for noisy, n=4 for pulsed, n=5 for tonal) were

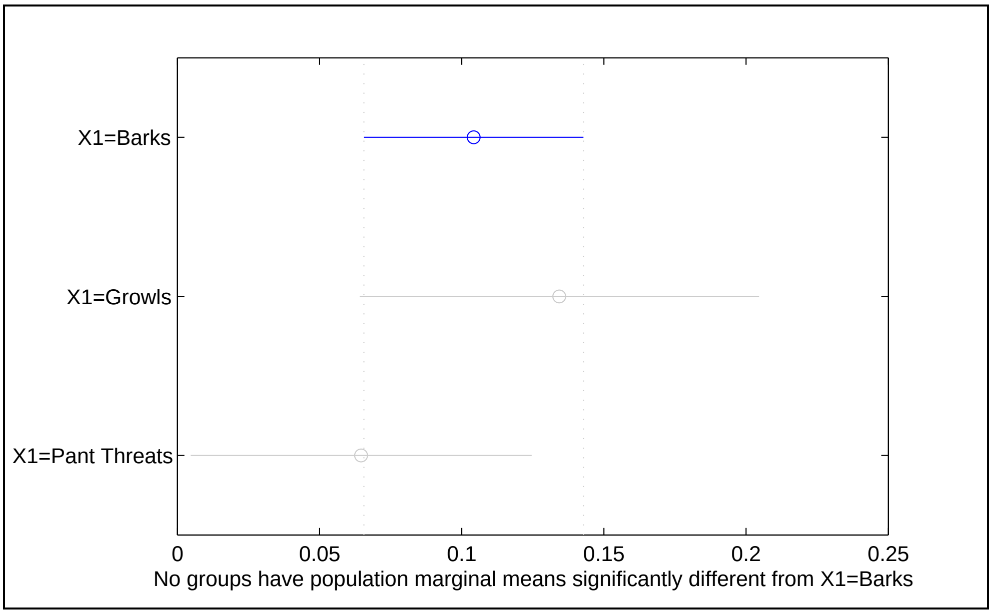


Figure 4.6 Marginal difference between aggressive subclasses

compared by using unbalanced one-way ANOVA test, which showed that the difference between subgroup means is significant (p<0.01). Resulting statistics were analyzed multivariate post-hoc test. Marginal differences among scream subtypes are shown in Figure 4.7

Multivariate data analysis revealed that arched scream mean is significantly different than noisy and pulsed screams, while noisy screams also differ from tonal screams. Mean and variance of LNM and mean of the maximum SNR results are given in Table 4.5.

Table 4.5
Mean and standard deviation of LNM, and mean of the maximum SNR for scream subtypes

	Mean LNM	Standard Deviation	Mean Maximum SNR
Arched	3.9931	1.1727	10.93
Noisy	0.2662	0.2926	1.72
Pulsed	1.5064	1.0088	5.74
Tonal	3.8030	1.5586	10.78

Current database did not include undulating screams; therefore it was not pos-

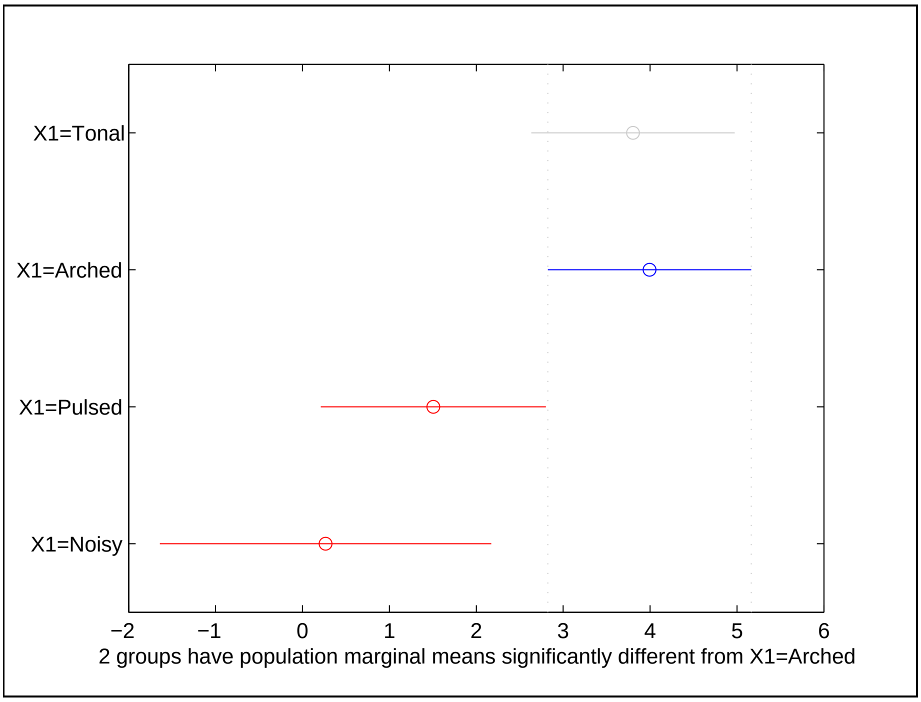


Figure 4.7 Marginal differences between scream subtypes

sible to compare this subtype with the others. However, there are two signals marked as "undulated>noisy"; and these signals had relatively higher LNM than noisy calls. Thus, it can be concluded that these calls are undulated screams.

4.6 Signal-to-Noise Ratio vs. Low-dimensional Nonlinearity

Low-dimensional nonlinearity measure for all signals was compared to maximum signal to noise ratio obtained for all percentage of neighbors. LNM seems to increase with SNR, but with linear regression, a weak correlation between SNR and LNM values is found. For all signals, slope of linear fit is 0.1670 and norm of residuals is 20.19. For signals with high LNM ($LNM>1$), this correlation was higher; the slope is 0.1206 and norm of residuals is 10.717. Correlation plots are shown in Figure 4.8 for all signals and in Figure 4.9 for signals with high LNM.

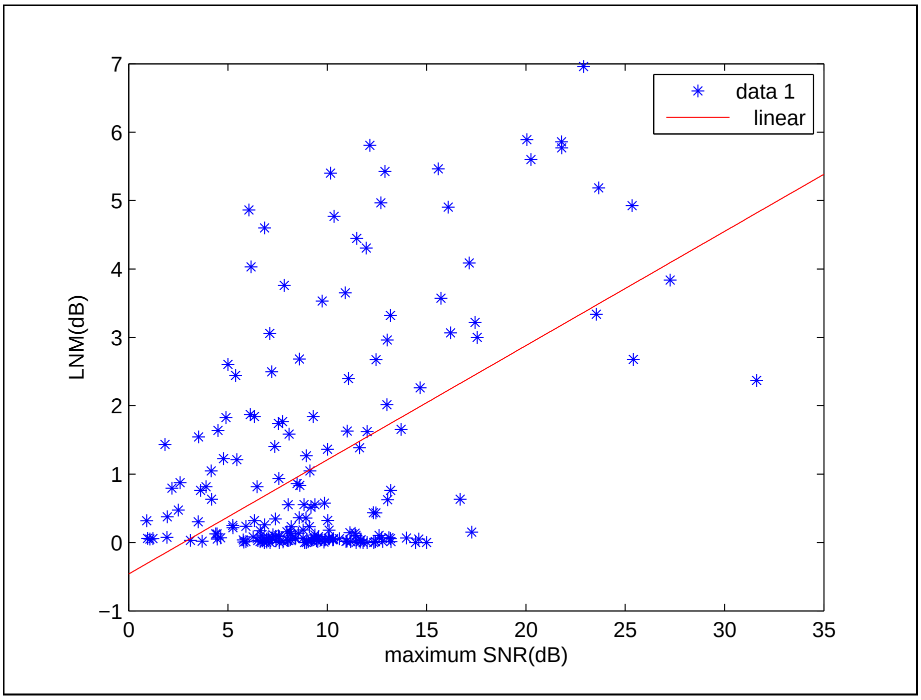


Figure 4.8 Correlation plot for all signals

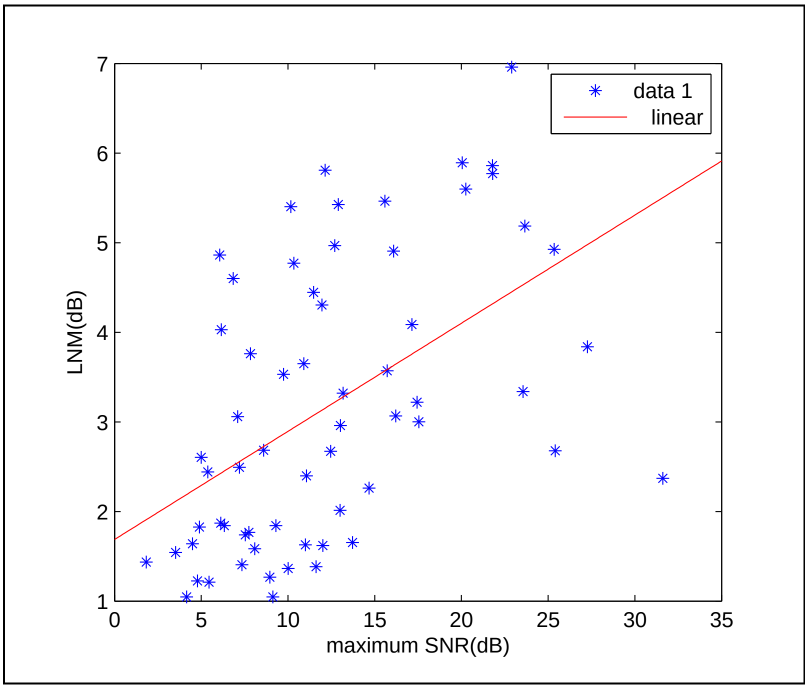


Figure 4.9 Correlation plot for signals which show low-dimensional nonlinearity

5. DISCUSSION

In this thesis, nonlinearity of monkey voice signals was studied by DVS analysis and this measure was examined among different call types for all animals, different call types of one single animal, coo calls of two different animals, pant threats and grunts, and subtypes of aggressive calls and screams.

For comparisons among different call types, both analysis for all animals and for a single animal, I found that coos, harmonic arches and screams have relatively higher LNM value and thus they show a higher low-dimensional nonlinearity than aggressive calls, grunts and girneys. High LNM value for coos and harmonic arches was expected, as in most of the cases, a fundamental frequency can be distinguished from frequency domain analysis of these signals. In some cases, coos can show chaotic behavior [3] , which also results in decreasing in number of dimensions, therefore in high LNM value. Similarly, low LNM mean of aggressive calls was also expected as these signals are highly irregular.

Screams are type of signals that are not stationary in the time course. What is surprising when we look from the conceptual point of view is that while screams are not ordinary communicative signals and in this sense their purpose should be to take attention of other individuals (in order to call for help etc.), resulting LNM was rather high in screams, especially for tonal and arched calls. In this sense, maybe it is not the irregularity of the signal which takes attention of other individuals; but the unstationarity may make the voice signal distinctive among other voices that other individuals hear.

Another important point about screams is that there are five different subtypes which differ not only in complexity and frequency, but also in context. We showed that nonlinearity measure can distinguish the difference between subtypes in most of cases. Scream signals are the most studied call types with coos because of their richness of context. In this sense of variety of contexts, it is somehow expected that screams and coos have high LNM value and high standard deviation. Nonlinearity analysis on human voices like crying, screaming, laughing may be useful to compare this phenomenon to human vocalizations.

Call types that correspond to social interactions resulted in moderate LNM value. Girneys and grunts are this type of call groups. Nevertheless, it should be noted that girney and grunt means were not significantly different from aggressive call mean therefore it is not possible to make a conclusion about the context by only looking at the nonlinearity measure.

Fitch et al. [3] proposed that the nonlinearity of the signal may change between animals and be useful for individual recognition. We applied a t-test to different coo samples from two animals in order to see if there is a significant difference of nonlinearity among calls of different individuals. The t-test did not reveal any significant difference between different callers. Nevertheless, it should be noted that the dataset that we analyzed for different call types of the same animal was too small, and therefore future analysis with more data may give more reliable results. In addition, comparison of nonlinearity of voices from different individuals require more detailed tests that investigate the bifurcations, fundamental frequency (or frequencies for voices with subharmonics), number of limit cycles and their properties etc.

It was reported in the documentation of vocal database that grunts are very similar to pant threat calls, but these two call types are given in very different contexts. We analyzed the difference between these two call types in order to see if the nonlinearity measure may distinguish a difference between these two call types. Two-sample t-test did not show any difference ($p>0.05$), so LNM measure can not be used to distinguish between these two call types. Similarly, ANOVA test between aggressive subgroups did not reveal any difference between group subtypes ($p>0.05$). These results show that although there is significance between mean differences for main call groups, the LNM is not very useful itself to distinguish similar calls with low LNM. The linear correlation between LNM and SNR also decreased for signals with low LNM. Thus, nonlinearity measure is not very useful to analyze conceptual information from the complex signals.

In addition, the weak correlation between SNR and LNM may be a result of low LNM obtained from the majority of the data in monkey voice database. Higher correlation for signals that show low-dimensional nonlinearity supports this idea.

6. CONCLUSION

In this study, rhesus macaque vocalizations were compared by using a nonlinear time series analsis method. Results were used to compare individual properties of animals.

Nonlinearity measure was useful to roughly distinguish the difference between call types. However, it fails to distinguish subtle properties of the voice that carries individual information. Except the scream subtype analysis, the results obtained by this method did not carry any different information than spectral analysis. Therefore, this method can not be a replacement to spectral analysis, but it may serve to understand the underlying dynamics behind the voice production. Moreover, scream subtype analysis shows that this method can be used with the spectrograms where spectral analysis methods cannot give any more information.

Another important point here is that the database was small for detailed analysis like individual recognition. Therefore, results for these small groups may not be reliable enough and more detailed analysis on larger databases may give better results.

One other important analysis on this dataset can be chaos tests, but as the voice data is very unstationary, Lyapunov exponent calculation and some other methods become more difficult. Another possible method can be to use Poincaré Maps on reconstructed phase space; but the problem with this method is the large noise term in the data. Thus, future work may be to investigate the underlying dynamics in a more detailed manner.

REFERENCES

1. Fitch, W., "The evolution of speech: a comparative review," *Trends in Cognitive Sciences*, Vol. 4, no. 7, pp. 258–267, 2000.
2. Sciamarella, D., and G. Mindlin, "Topological Structure of Chaotic Flows from Human Speech Data," *Physical Review Letters*, Vol. 82, no. 7, pp. 1450–1453, 1999.
3. Fitch, W., J. Neubauer, and H. Herzel, "Calls out of chaos: the adaptive significance of nonlinear phenomena in mammalian vocal production," *Animal Behaviour*, Vol. 63, no. 3, pp. 407–418, 2002.
4. Wilden, I., H. Herzel, G. Peters, and G. Tembrock, "Subharmonics, biphonation, and deterministic chaos in mammal vocalization," *Bioacoustics*, Vol. 9, no. 3, pp. 171–196, 1998.
5. Hauser, M., *The Evolution of Communication*, MIT Press, 1996.
6. Berry, D., H. Herzel, I. Titze, and K. Krischer, "Interpretation of biomechanical simulations of normal and chaotic vocal fold oscillations with empirical eigenfunctions," *The Journal of the Acoustical Society of America*, Vol. 95, p. 3595, 1994.
7. Herzel, H., D. Berry, I. Titze, and I. Steinecke, "Nonlinear dynamics of the voice: Signal analysis and biomechanical modeling," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, Vol. 5, p. 30, 1995.
8. Steinecke, I., and H. Herzel, "Bifurcations in an asymmetric vocal-fold model," *The Journal of the Acoustical Society of America*, Vol. 97, p. 1874, 1995.
9. Tokuda, I., T. Riede, J. Neubauer, M. Owren, and H. Herzel, "Nonlinear analysis of irregular animal vocalizations," *The Journal of the Acoustical Society of America*, Vol. 111, p. 2908, 2002.
10. Ruelle, D., and F. Takens, "On the nature of turbulence," *Communications in Mathematical Physics*, Vol. 20, no. 3, pp. 167–192, 1971.
11. Takens, F., *et al.*, "Detecting strange attractors in turbulence," *Lecture Notes in Mathematics*, Vol. 898, no. 1, pp. 366–381, 1981.
12. Broomhead, D., R. Indik, A. Newell, and D. Rand, "Local adaptive Galerkin bases for large-dimensional dynamical systems," *Nonlinearity*, Vol. 4, pp. 159–197, 1991.
13. Casdagli, M., "Chaos and deterministic versus stochastic nonlinear modelling," *Journal of the Royal Statistical Society B*, Vol. 54, no. 2, pp. 303–328, 1991.
14. Miller, C., and A. Ghazanfar, "Meaningful acoustic units in nonhuman primate vocal behavior," *The Cognitive Animal*, pp. 265–273, 2002.
15. Website, T. V. P., 2004. Available: http://www.voiceproblem.org/.
16. Hauser, M., C. Evans, and P. Marler, "The role of articulation in the production of rhesus monkey (Macaca mulatta) vocalizations," *Anim. Behav*, Vol. 45, pp. 423–433, 1993.
17. Gouzoules, S., H. Gouzoules, and P. Marler, "Rhesus monkey (Macaca mulatta) screams: Representational signalling in the recruitment of agonistic aid," *Animal Behaviour*, Vol. 32, no. 1, pp. 182–193, 1984.

18. Farmer, J., and J. Sidorowich, "Predicting chaotic time series," *Physical Review Letters*, Vol. 59, no. 8, pp. 845–848, 1987.

19. Markel, J., and A. Gray, *Linear Prediction of Speech*, Springer-Verlag New York, Inc. Secaucus, NJ, USA, 1982.

20. Abarbanel, H., R. Brown, J. Sidorowich, and L. Tsimring, "The analysis of observed chaotic data in physical systems," *Reviews of Modern Physics*, Vol. 65, no. 4, pp. 1331–1392, 1993.

21. Casdagli, M., S. Eubank, J. Farmer, and J. Gibson, "State space reconstruction in the presence of noise," *Physica D*, Vol. 51, no. 1-3, pp. 52–98, 1991.